

Relationships among some chi-square
goodness of fit statistics

BU-843-M

Revised
October, 1984

by

Charles E. McCulloch

RELATIONSHIPS AMONG SOME CHI-SQUARE
GOODNESS OF FIT STATISTICS

Charles E. McCulloch

Cornell University
Ithaca, N.Y.

*Key Words and Phrases: Rao-Robson statistic; Dzhaparidze-Nikulin
statistic; maximum likelihood*

ABSTRACT

Relationships among the Rao-Robson, Dzhaparidze-Nikulin, and Pearson-Fisher chi-square goodness of fit statistics are established. These statistics are also shown to be related to a statistic which compares the maximum likelihood estimator based on the original observations with the maximum likelihood estimator based on the cell frequencies and a statistic proposed by Hsuan. These relationships are used as a basis to understand the performance of the statistics and to suggest a new goodness of fit statistic.

1. INTRODUCTION

In this paper we deal with relationships among several chi-square goodness of fit statistics. We will consider the setting in which the data are grouped into cells and the objective is to test goodness of fit to a distribution with unspecified nuisance parameters, e.g., testing goodness of fit to a normal distribution with unknown mean and variance. The Rao-Robson statistic

(Rao, 1971; Rao and Robson, 1974) is shown to consist of two independent pieces, the Dzhaparidze-Nikulín statistic (Dzhaparidze and Nikulín, 1974) plus a statistic which behaves locally like a comparison of the maximum likelihood estimator based on the original observations with the maximum likelihood estimator based on the cell frequencies. We also show that the Pearson-Fisher statistic is asymptotically equivalent to the Dzhaparidze-Nikulín statistic under local alternatives. Connections are additionally made to a statistic proposed by Hsuan (1974) for testing fit to a normal distribution. These relationships are used to gain insight into the performance of the goodness of fit statistics and to propose a new statistic for testing fit.

2. NOTATION

Let $X_1 \sim \text{i.i.d. } G(x; \theta, \eta)$ ($1 \leq i \leq n$) with density or mass function $g(x; \theta, \eta)$, where $\theta = (\theta_1, \theta_2, \dots, \theta_s)$ is an unknown nuisance parameter and $\eta = (\eta_1, \eta_2, \dots, \eta_r)$ is an unknown parameter specifying deviations from the null distribution. θ_0 and η_0 will denote the true but unknown parameter values under the null hypothesis. For notational convenience, the parameters will often be omitted when they are equal to θ_0 or η_0 . Thus, $G(x; \theta, \eta_0) = G(x; \theta)$ is the family of distributions to which we are testing fit. Formally, we will be interested in testing $H_0: X_1 \sim \text{i.i.d. } G(x; \theta)$ versus $H_A: X_1 \sim \text{i.i.d. } F(x)$, where $F(\cdot) \neq G(\cdot; \theta)$ for all θ . The chi-square tests considered are based on dividing the sample space into cells, C_j ($1 \leq j \leq k$), and using the frequency of observations, f_j , falling in cell C_j . To construct the tests, a vector of standardized cell frequencies is used:

$$V(\theta, \eta) = \left(\frac{f_1 - np_1(\theta, \eta)}{[np_1(\theta, \eta)]^{\frac{1}{2}}}, \dots, \frac{f_k - np_k(\theta, \eta)}{[np_k(\theta, \eta)]^{\frac{1}{2}}} \right),$$

where $p_j(\theta, \eta)$ = probability of an observation falling in cell C_j

$$= \int_{C_j} dG(x; \theta, \eta) \quad .$$

When both the original data X_1, X_2, \dots, X_n and the cell frequencies, f_1, f_2, \dots, f_k , are available, two estimators of θ can be formed. Let $\hat{\theta}_x$ denote the estimator found by maximizing the likelihood derived from the original observations, $\prod_{i=1}^n g(x_i; \theta)$, and let $\tilde{\theta}_f$ be the maximum likelihood estimator for θ based on the likelihood from the cell frequencies, $\prod_{j=1}^k p_j(\theta)^{f_j}$.

Two other quantities are needed in the construction of the chi-square statistics:

$J(\theta)$ = information matrix for estimating θ from X_1, X_2, \dots, X_n

$$= \left[E_{\theta} \left(\frac{\partial \log g(X; \theta)}{\partial \theta_i} \cdot \frac{\partial \log g(X; \theta)}{\partial \theta_j} \right) \right]$$

and

$$B(\theta) = \left[p_i(\theta)^{-1} \frac{\partial p_i(\theta)}{\partial \theta_j} \right] \quad .$$

Note that $B(\theta)'B(\theta)$ is the information matrix for estimating θ from f_1, f_2, \dots, f_k .

3. CHI-SQUARE GOODNESS OF FIT STATISTICS

Using the notation in Section 2, the traditional, Pearson-Fisher goodness of fit statistic, PF, can be written as

$$PF = \sum_{j=1}^k \frac{\left(f_j - np_j(\tilde{\theta}_f) \right)^2}{np_j(\tilde{\theta}_f)} = V(\tilde{\theta}_f)' V(\tilde{\theta}_f) \quad .$$

This has an asymptotic χ^2_{k-s-1} distribution (s is the dimension of θ) under the null hypotheses but can be awkward to calculate since $\tilde{\theta}_f$ often does not have a closed form representation. Two other goodness of fit statistics which use different estimators

of θ , the Rao-Robson (RR) and Dzhaparidze-Nikulin (DN) statistics, can also be easily written:

$$RR = V(\hat{\theta}_x)' (I - BJ^{-1}B')^{-1} V(\hat{\theta}_x)$$

$$DN(\bar{\theta}) = V(\bar{\theta})' \left(I - B(B'B)^{-1}B' \right)^{-1} V(\bar{\theta}),$$

where

$$\bar{\theta} = \text{any estimator of } \theta \text{ with} \\ \text{order } n^{-\frac{1}{2}} \text{ convergence}$$

$$= \theta_0 + o_p(n^{-\frac{1}{2}}) \quad .$$

The statistic RR is essentially

$$V(\hat{\theta}_x)' \left(\text{Var}[V(\hat{\theta}_x)] \right)^{-1} V(\hat{\theta}_x) \quad ,$$

i.e., a standardized quadratic form in $V(\hat{\theta}_x)$. It has an asymptotic χ^2_{k-1} distribution under the null hypothesis. The statistic $DN(\bar{\theta})$ has an asymptotic χ^2_{k-s-1} distribution. In a sense to be made more precise, the Rao-Robson statistic recovers the information lost in estimating $\theta_1, \theta_2, \dots, \theta_s$.

In the section that follows, we will need one step asymptotic approximations for $\hat{\theta}_x, \tilde{\theta}_f$, and $V(\bar{\theta})$ under local alternatives. We consider local alternatives parameterized by $\eta_n = \eta_0 + n^{-\frac{1}{2}}\gamma$, where γ is fixed. Under mild regularity conditions (see Moore and Spruill, 1975 for details) the following expansions hold

$$\begin{aligned} \hat{\theta}_x &= \theta_0 + n^{-1} J^{-1} \left(\sum_{i=1}^n \frac{\partial \log g(X_i, \theta_0, \eta_n)}{\partial \theta} \right) \\ &\quad + n^{-\frac{1}{2}} J^{-1} J_{12} \gamma + o_p(n^{-\frac{1}{2}}) \quad , \end{aligned} \quad (3.1)$$

$$\tilde{\theta}_f = \theta_0 + n^{-\frac{1}{2}} (B'B)^{-1} B' V(\eta_n) + n^{-\frac{1}{2}} (B'B)^{-1} B' B_{12} \gamma + o_p(n^{-\frac{1}{2}}), \quad (3.2)$$

$$V(\bar{\theta}) = V(\eta_n) - B n^{\frac{1}{2}} (\bar{\theta} - \theta_0) + B_{12} \gamma + o_p(1) \quad , \quad (3.3)$$

where

$$J_{12} = \left(E \left[\frac{\partial \log g}{\partial \theta_1} \cdot \frac{\partial \log g}{\partial \eta_j} \right] \right) ,$$

$$B_{12} = \left(p_1(\theta_0)^{-\frac{1}{2}} \frac{\partial p_1(\theta_0)}{\partial \eta_j} \right) ,$$

and $\bar{\theta}$ is any estimator of θ .

4. RELATIONSHIPS

In this section we derive relationships among the chi-square statistics. First we derive an asymptotic local expression for

$$n(\hat{\theta}_x - \tilde{\theta}_f)' \left[\text{Var}(\hat{\theta}_x - \tilde{\theta}_f) \right]^{-1} (\hat{\theta}_x - \tilde{\theta}_f).$$

Theorem 4.1 Under (θ_0, η_n) and assuming (3.1) - (3.3) are valid,

$$\begin{aligned} & n(\hat{\theta}_x - \tilde{\theta}_f)' \left[\text{Var}(\hat{\theta}_x - \tilde{\theta}_f) \right]^{-1} (\hat{\theta}_x - \tilde{\theta}_f) \\ &= n(\hat{\theta}_x - \tilde{\theta}_f)' \left((B'B)^{-1} - J^{-1} \right)^{-1} (\hat{\theta}_x - \tilde{\theta}_f) + o_p(1) \\ &= V(\hat{\theta}_x)' B \left[(J - B'B)^{-1} + (B'B)^{-1} \right] B' V(\hat{\theta}_x) + o_p(1). \quad (4.1) \end{aligned}$$

Proof: To see the first identity, the asymptotic variance of $n^{\frac{1}{2}}(\hat{\theta}_x - \tilde{\theta}_f)$ under (θ_0, η_0) can be calculated from (3.1) and (3.2):

$$\text{asymptotic Var}(n^{\frac{1}{2}}\hat{\theta}_x) = J^{-1}$$

$$\text{asymptotic Var}(n^{\frac{1}{2}}\tilde{\theta}_f) = (B'B)^{-1}$$

$$\text{asymptotic Cov}(n^{\frac{1}{2}}\hat{\theta}_x, n^{\frac{1}{2}}\tilde{\theta}_f) = J^{-1} .$$

Thus,

$$\begin{aligned} \text{asymptotic Var} \left(n^{\frac{1}{2}}(\hat{\theta}_x - \tilde{\theta}_f) \right) &= J^{-1} + (B'B)^{-1} - 2J^{-1} \\ &= (B'B)^{-1} - J^{-1} . \end{aligned}$$

To prove the second identity, solve identity (3.3) for $V(\eta_n)$ and use $\bar{\theta} = \hat{\theta}_x$. This gives the expansion

$$V(\eta_n) = V(\hat{\theta}_x) + Bn^{\frac{1}{2}}(\hat{\theta}_x - \theta_0) - B_{12}\gamma + o_p(1) .$$

Plugging this into expansion (3.2) gives the formula

$$\begin{aligned} \tilde{\theta}_f &= \theta_0 + n^{-\frac{1}{2}}(B'B)^{-1}B' \left[V(\hat{\theta}_x) + Bn^{\frac{1}{2}}(\hat{\theta}_x - \theta_0) - B_{12}\gamma + o_p(1) \right] \\ &\quad + n^{-\frac{1}{2}}(B'B)^{-1}B'B_{12}\gamma + o_p(n^{-\frac{1}{2}}) \\ &= \theta_0 + n^{-\frac{1}{2}}(B'B)^{-1}B'V(\hat{\theta}_x) + \hat{\theta}_x - \theta_0 - n^{-\frac{1}{2}}(B'B)^{-1}B'B_{12}\gamma \\ &\quad + n^{-\frac{1}{2}}(B'B)^{-1}B'B_{12}\gamma + o_p(n^{-\frac{1}{2}}) \\ &= n^{-\frac{1}{2}}(B'B)^{-1}B'V(\hat{\theta}_x) + \hat{\theta}_x + o_p(n^{-\frac{1}{2}}) . \end{aligned} \tag{4.1}$$

Rearranging (4.1) gives

$$n^{\frac{1}{2}}(\tilde{\theta}_f - \hat{\theta}_x) = (B'B)^{-1}B'V(\hat{\theta}_x) + o_p(1) . \tag{4.2}$$

Using (4.2) in the quadratic form yields the following

$$\begin{aligned} &n(\tilde{\theta}_f - \hat{\theta}_x)' \left((B'B)^{-1} - J^{-1} \right)^{-1} (\tilde{\theta}_f - \hat{\theta}_x) \\ &= V(\hat{\theta}_x)' B(B'B)^{-1} \left((B'B)^{-1} - J^{-1} \right)^{-1} (B'B)^{-1} B'V(\hat{\theta}_x) + o_p(1) \\ &= V(\hat{\theta}_x)' B(J - B'B)^{-1} J(B'B)^{-1} B'V(\hat{\theta}_x) + o_p(1) \end{aligned}$$

$$= V(\hat{\theta}_x)' B \left((J - B'B)^{-1} + (B'B)^{-1} \right) B' V(\hat{\theta}_x) + o_p(1) \quad \parallel$$

We are now in a position to derive the first correspondence between goodness of fit statistics. Using the identity

$$(I - BJ^{-1}B')^{-1} = I - B(B'B)^{-1}B' + B \left((J - B'B)^{-1} + (B'B)^{-1} \right) B',$$

we can partition RR into two pieces:

$$\begin{aligned} RR &= V(\hat{\theta}_x)' (I - BJ^{-1}B')^{-1} V(\hat{\theta}_x) \\ &= V(\hat{\theta}_x)' \left(I - B(B'B)^{-1}B' \right) V(\hat{\theta}_x) + \\ &\quad V(\hat{\theta}_x)' B \left((J - B'B)^{-1} + (B'B)^{-1} \right) B' V(\hat{\theta}_x) \\ &= DN(\hat{\theta}_x) + V(\hat{\theta}_x)' B \left((J - B'B)^{-1} + (B'B)^{-1} \right) B' V(\hat{\theta}_x) \quad (4.2) \end{aligned}$$

This partitioning shows that the Rao-Robson statistic consists of two pieces: the Dzhaparidze-Nikulín statistic and the statistic derived in Theorem 4.1. Further, under the null hypothesis, these two pieces are asymptotically independent, each with χ^2 distributions. These facts are summarized in Theorem 4.2.

Theorem 4.2 Under (θ_0, η_n) and assuming (3.1) - (3.3) are valid,

$$\begin{aligned} RR &= DN(\hat{\theta}_x) + V(\hat{\theta}_x)' B \left((J - B'B)^{-1} + (B'B)^{-1} \right) B' V(\hat{\theta}_x) \\ &= DN(\hat{\theta}_x) + n(\hat{\theta}_x - \tilde{\theta}_f)' \left((B'B)^{-1} - J^{-1} \right)^{-1} (\hat{\theta}_x - \tilde{\theta}_f) + o_p(1) \quad . \end{aligned}$$

Further, under (θ_0, η_0) ,

(a) $DN(\hat{\theta}_x)$ is distributed asymptotically χ^2_{k-s-1} and

(b) $V(\hat{\theta}_x)' B \left((J - B'B)^{-1} + (B'B)^{-1} \right) B' V(\hat{\theta}_x)$ or,

equivalently, $n(\hat{\theta}_x - \tilde{\theta}_f) \left((B'B)^{-1} - J^{-1} \right)^{-1} (\hat{\theta}_x - \tilde{\theta}_f)$,

has an asymptotic χ^2_g distribution, independent of $DN(\hat{\theta}_x)$.

Proof: The first two assertions are proved above in (4.1) and (4.2). The asymptotic distributions follow from the usual theory on quadratic forms in normal variates (Moore, 1976) since

$$V(\hat{\theta}_x) \sim AN(0, I - BJ^{-1}B') ,$$

where $AN(\cdot, \cdot)$ denotes asymptotic normality. For example, for independence we need to check that

$$\left(I - B(B'B)^{-1}B' \right) \left(I - BJ^{-1}B' \right) \left(B \left[(J - B'B)^{-1} + (B'B)^{-1} \right] B' \right)$$

is zero, which is easily seen to be true. \parallel

The next relationship that we derive shows that $DN(\hat{\theta}_x)$ behaves locally like the Pearson-Fisher statistic.

Theorem 4.3 Under (θ_0, η_n) and assuming (3.1) - (3.3) hold,

$$DN(\bar{\theta}) = PF + o_p(1) .$$

Proof: Using (3.3) with $\bar{\theta} = \tilde{\theta}_f$ gives

$$V(\tilde{\theta}_f) = V(\eta_n) - Bn^{\frac{1}{2}}(\tilde{\theta}_f - \theta_0) + B_{12}\gamma + o_p(1) .$$

Plugging in (3.2) yields the expression

$$\begin{aligned} V(\tilde{\theta}_f) &= V(\eta_n) - B \left((B'B)^{-1}B'V(\eta_n) + (B'B)^{-1}B'B_{12}\gamma \right) + B_{12}\gamma + o_p(1) \\ &= \left(I - B(B'B)^{-1}B' \right) V(\eta_n) + \left(I - B(B'B)^{-1}B' \right) B_{12}\gamma + o_p(1) \\ &= \left(I - B(B'B)^{-1}B' \right) \left(V(\eta_n) - Bn^{\frac{1}{2}}(\bar{\theta} - \theta_0) + B_{12}\gamma \right) + o_p(1) \end{aligned}$$

$$= \left(I - B(B'B)^{-1}B' \right) V(\bar{\theta}) + o_p(1) \quad .$$

This result shows that under (θ_0, η_n) ,

$$\begin{aligned} DN(\bar{\theta}) &= V(\bar{\theta})' \left(I - B(B'B)^{-1}B' \right) \left(I - B(B'B)^{-1}B' \right) V(\bar{\theta}) \\ &= V(\tilde{\theta}_f)' V(\tilde{\theta}_f) + o_p(1) \\ &= PF + o_p(1) \quad \parallel \end{aligned}$$

Corollary 4.4 Under (θ_0, η_n) and assuming (3.1) - (3.3) hold

$$RR = PF + n(\hat{\theta}_x - \tilde{\theta}_f)' \left((B'B)^{-1} - J^{-1} \right)^{-1} (\hat{\theta}_x - \tilde{\theta}_f) + o_p(1)$$

Proof: combine Theorem 4.3 and Theorem 4.2 ||

5. DISCUSSION

The preceding results lead to an interesting interpretation of the Rao-Robson statistic. The interpretation is that it recovers (locally) degrees of freedom lost in estimating θ from grouped data. The specific form that this recovery of information takes is a standardized comparison of $\hat{\theta}_x$ and $\tilde{\theta}_f$.

Monte Carlo simulations by Rao and Robson (1974), Hsuan (1974), and McCulloch (1980) indicate a significant increase in power for the Rao-Robson statistic over the Pearson-Fisher or Dzharidze-Nikulín statistic and this suggests the Rao-Robson statistics should be used whenever ungrouped data are available. This dramatic increase in power also suggests that the statistic

$$n(\hat{\theta}_x - \tilde{\theta}_f)' \left((B'B)^{-1} - J^{-1} \right)^{-1} (\hat{\theta}_x - \tilde{\theta}_f)$$

or its local equivalent,

$$V(\hat{\theta}_x)' B \left((J - B'B)^{-1} + (B'B)^{-1} \right) B' V(\hat{\theta}_x) \quad (5.1)$$

may be a good test statistic in its own right. In fact, the statistic proposed by Hsuan (1974) for testing fit to univariate normality can be shown (McCulloch, 1980) to be a random cell version of

$$V(\hat{\theta}_x)'B \left((J - B'B)^{-1} + (B'B)^{-1} \right) B'V(\hat{\theta}_x) .$$

McCulloch (1980) also investigated random cell versions for testing fit to a multivariate normal distribution. In a simulation evaluating the power of tests of fit to a bivariate normal distribution, the Rao-Robson statistic and the statistic in (5.1) performed quite well. The Rao-Robson statistic performed on par with multivariate generalizations of powerful univariate tests of fit as considered in Malkovich and Afifi (1973). The statistic in (5.1) did not perform quite as well overall, but still outperformed the other tests for some alternatives. Alternatives considered included contaminated bivariate normals, distributions with heavier tails than the bivariate normal and non-elliptic distributions.

These results run counter to the reasoning of Watson and Kendall and Stuart that the statistic (5.1) should be ignored. Watson (1959) states that, "...the obvious practical method is to use as many class intervals as possible so that the disturbance term may be ignored altogether." Kendall and Stuart (1973) echo this sentiment, "...as k becomes large these $(\hat{\theta}_x$ and $\tilde{\theta}_f)$ are so close together that the difference can be ignored."

Random cell versions of these statistics, where the cells depend on consistent estimators of the unknown parameters, will have the same relationships as outlined in this article. Moore and Spruill (1975) show that as long as the random cells converge to some limit in probability, they have no effect on the asymptotic distributions.

ACKNOWLEDGEMENT

The author thanks D.S. Robson for guidance on the Ph.D. thesis on which this article is based.

BIBLIOGRAPHY

- Dzhaparidze, K. O. and Nikulin, M. S. (1974). On a modification of the standard statistics of Pearson. *Theory of Probability and Its Applications* 19: 851-853.
- Hsuan, T. A. (1974). Some chi-square statistics for goodness of fit tests. Ph.D. thesis, Cornell University, Ithaca, New York.
- Kendall, M. G. and Stuart, A. (1977). *The Advanced Theory of Statistics Vol. 2* MacMillan Publishing Co., Inc., New York.
- Malkovich, J. F. and Afifi, A. A. (1973). On tests for multivariate normality. *Journal of the American Statistical Association* 68: 176-179.
- McCulloch, C. E. (1980). Construction of conditional statistics for chi-square goodness of fit tests. Ph.D. thesis, Cornell University, Ithaca, New York.
- Moore, D. S. (1976). Recent developments in chi-square tests for goodness of fit. Department of Statistics Mimeograph Series #459, Purdue University, Lafayette, Indiana.
- Moore, D. S. and Spruill, M. C. (1975). Unified large-sample theory of general chi-squared statistics for tests of fit. *Ann. Statist.* 3: 599-616.
- Rao, K. C. (1971). A chi-square statistic for goodness-of-fit tests. Ph.D. thesis, Cornell University, Ithaca, New York.
- Rao, K. C. and Robson, D. S. (1974). A chi-square statistic for goodness-of-fit tests within the exponential family. *Commun. Statist.* 3: 1139-1153.
- Watson, G. S. (1959). Some recent results in chi-square goodness-of-fit tests. *Biometrics* 15: 440-468.